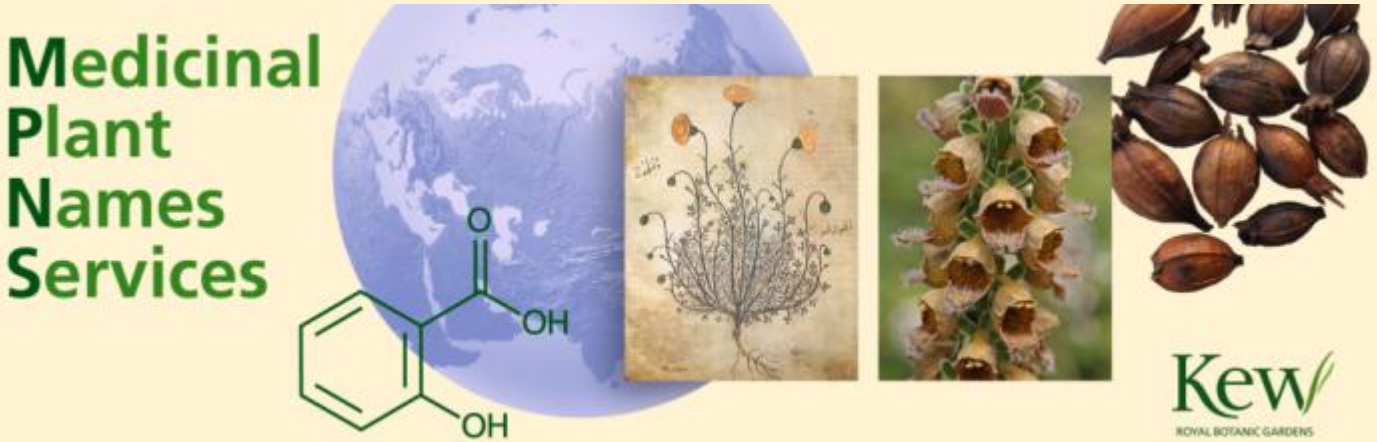


Medicinal Plant Names Services

Structure of Presentation

1 Introduce MPNS

2 What is a scientific plant name?

3 Issues when using scientific plant names

4 Impact & relevance

5 Global plant name resources

6 Role of MPNS and relevance to IDMP

Medicinal Plant Names Services

What is MPNS?

Information services designed for health, pharmacovigilance, research

Scope:

- 'medicinal plants':
  prioritising those in trade
- authoritative scientific plant nomenclature:
  comprehensive, integrated with & maintained by botanical community
- names as they are used: in pharmacopoeia, databases, legislation, etc
  map common names & herbal ingredients onto plant names
- reference monographs, uses, plant parts
- be part of information landscape: linking to partners

Structure of Presentation

1 Introduce MPNS

2 What is a scientific plant name?

3 Issues when using scientific plant names

4 Impact & relevance

5 Global plant name resources

6 Role of MPNS and relevance to IDMP

What is a Scientific Plant Name?

- Written in Latin
- Genus + species + author

“***Hocus pocus*** Bob” X

- International Nomenclatural Code for Botany:
  - establish protocols defining “formal publication”
  - reviewed every 5 years
  - **Author must**
    - state why this plant is different
    - cite the physical specimens seen (& used for description)
    - provide unique identifier and location of those specimens
    - E.g. ‘Polhill & Paulo #2,139 (1962) Kew,Paris,Nairobi,Missouri’

= “TYPE” specimens

Why we need scientific plant names?

1. Each is unique
   a) Formally published: nomenclatural Code
   b) Registered: International Plant Name Index
   c) Establish identity of the species: one dimension of formal description of a herbal ingredient

2. The meaning of each published name is fixed:
   a) Type specimens offer a reference point
   b) Can resolve disputes
   c) Meaning cannot change over time

3. International
   a) in language
   b) plants cross national boundaries
   c) traded globally
   d) research published worldwide

4. Relevance to legislation (e.g.)
   a) CITES: (UN CBD) Conservation & Customs restrictions
   b) Food & Agriculture: labelling
   c) Underlie DNA barcoding

Structure of Presentation

1 Introduce MPNS

2 What is a scientific plant name?

3 Issues when using scientific plant names

4 Impact & relevance

5 Global plant name resources

6 Role of MPNS and relevance to IDMP

Nevertheless…..
Issues using Scientific Plant Names

1. **Synonyms: too many names**

   1.6 million unique names [International Plant Name Index www.ipni.org]

   c. 0.4 million plants [The Plant List www.tpl.org]

   i.e. each plant (species) has multiple names

   important information published under any 1 of those names

2. **Homonyms: ambiguity**

   >5 % binomials published by >2 authors

   i.e. name may refer to > 1 plant (species) – 'author' is not optional!

3. **Misapplied names: inappropriate use**

   documented cases of regulatory, pharmacopoeia & research literature citing chemistry/use of one plant whilst using name of another

   i.e. Need to record, flag and divert users from false conclusions

Issues using Scientific Plant Names (cont)

4. **Names change: c.10,000 changes published /yr**
   - New species found (2K)
   - New data results in:
     - species moving into new genera (4K)
     - species being split or multiple species being merged (4K)

5. **Lack of a complete authoritative reference**
   - Synonymy information scattered in diverse resources
   - No authoritative list covers all medicinal plants
   - Existing resources designed for biologists
     - Unintelligible to health / pharmacovigilance community

Structure of Presentation

1 Introduce MPNS

2 What is a scientific plant name?

3 Issues when using scientific plant names

4 Impact relevance

5 Global plant name resources

6 Role of MPNS and relevance to IDMP

Impact of these “issues”?

- When looking for information
- When communicating
- For the regulatory process
- Implications for data standards/ structures

1) Impact when searching for information

a) User fails to find data

stored under mispelt names

stored under alternative synonyms

b) User derives the wrong conclusions

Homonyms: false assumptions

Misapplied names: erroneous data

Out of date taxonomy: misleading chemical patterns

c) Time & Effort

where do you find synonyms? check names?

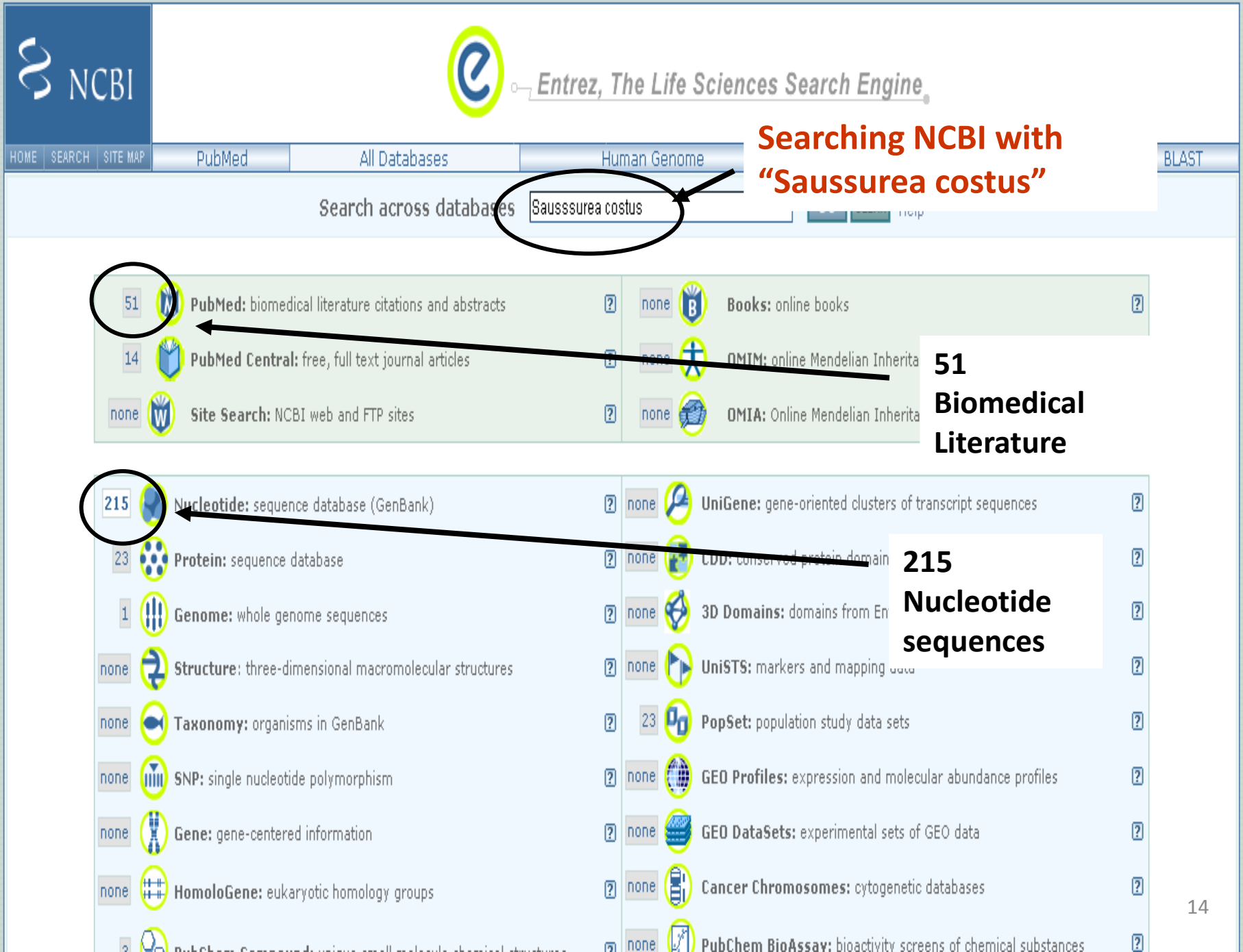
EXAMPLE 1: ‘Mu Xiang’

widely used in
Chinese Herbal Medicine

“Saussurea costus (Falc.) Lipschitz”

- Each national pharmacopoeia uses different name
- Chinese Pharmacopoeia calls this plant:
  “***Aucklandia lappa*** Decne”

- > 20 other scientific synonyms


NCBI
Entrez, The Life Sciences Search Engine
HOME
SEARCH
SITE MAP
PubMed
All Databases
Human Genome
BLAST
Search across databases
Sausssurea costus
Searching NCBI with "Saussurea costus"
51 PubMed: biomedical literature citations and abstracts
14 PubMed Central: free, full text journal articles
none Site Search: NCBI web and FTP sites
none Books: online books
OMIM: online Mendelian Inherita
none OMIA: Online Mendelian Inherita
51 Biomedical Literature
215 Nucleotide: sequence database (GenBank)
23 Protein: sequence database
1 Genome: whole genome sequences
none Structure: three-dimensional macromolecular structures
none Taxonomy: organisms in GenBank
none SNP: single nucleotide polymorphism
none Gene: gene-centered information
none HomoloGene: eukaryotic homology groups
none UniGene: gene-oriented clusters of transcript sequences
none 3D Domains: domains from En
none UniSTS: markers and mapping
23 PopSet: population study data sets
none GEO Profiles: expression and molecular abundance profiles
none GEO DataSets: experimental sets of GEO data
none Cancer Chromosomes: cytogenetic databases
none PubChem BioAssay: bioactivity screens of chemical substances
215 Nucleotide sequences


NCBI
Entrez, The Life Sciences Search Engine
HOME
SEARCH
SITE MAP
PubMed
All Databases
Human Genome
BLAST
Search across databases
Aucklandia lappa
Searching NCBI with "Aucklandia lappa"
3 PubMed: biomedical literature citations and abstracts
9 PubMed Central: free, full text journal articles
none Site Search: NCBI web and FTP sites
none Books: online books
none OMIM: online Mendelian I
none OMIA: Online Mendelian I
03 Biomedical Literature
14 Nucleotide: sequence database (GenBank)
5 Protein: sequence database
none Genome: whole genome sequences
1 Structure: three-dimensional macromolecular structures
none Taxonomy: organisms in GenBank
none SNP: single nucleotide polymorphism
none Gene: gene-centered information
none HomoloGene: eukaryotic homology groups
none PubChem Compound: unique small molecule chemical structures
none UniGene: gene-oriented clusters of transcript sequences
none CDD: conserved protein domain
3 3D Domains: domains from En
none UniSTS: markers and mapping
8 PopSet: population study data sets
none GEO Profiles: expression and molecular abundance profiles
none GEO DataSets: experimental sets of GEO data
none Cancer Chromosomes: cytogenetic databases
none PubChem BioAssay: bioactivity screens of chemical substances
14 Nucleotide sequences

IMPACT on users of NCBI?

- Users fail to locate information

  a) users will not know of > 20 synonyms

     c. 25% of NCBI records relate to synonyms only

  b) 10% data records are stored in NCBI under misspelt names

     records may never be found

- Users draw wrong conclusions:

     c. 5% of NCBI records use homonyms (to >1 species) ambiguous

     User could for example assume chemical properties for wrong plant

2) Impact when publishing information

a) Your information becomes untraceable

Did your publish data under mispelt or made-up name?
Did you use an older synonym?

b) You may mislead or be ambiguous

Did you list the same plant > 1 under different synonyms?
Does the name used potentially refer to >1 species?
Are you implying chemical similarities that do not exist?

c) Time & Effort

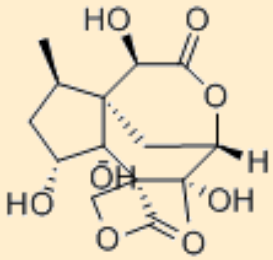
how do you ensure the integrity of your publication?

Example 2 Star Anise

EU Commission
Decision #OJEC L 2.2.2002 L 33/31:

*"the botanical variety **Illicium anisatum** is scientifically recognised as highly poisonous"*

in doing so they
potentially banned the import of Star Anise!



‘Japanese Star Anise’ (Toxic) ***Illicium anisatum*** Linnaeus

commonly used synonyms *Illicium religiosum* S&Z.

Bandianifera anisatum Kuntze

Ambiguity !!!!!

‘Star Anise’ ***Illicium verum*** Hook.f.

commonly used synonyms *Illicium anisatum* Lour.

Illicium san-ki Perr.

Bandianifera officianarum Kuntze

Two further published homonyms

Illicium anisatum Bartr. ex. Michx

Illicium anisatum Gaertn.

3) Impact on regulatory process

a) How might you establish whether a plant name has been formally published (i.e. is scientifically meaningful)?

- need access to ALL published plant names!
- a list of medicinal plants isn't good enough

b) How to ensure coherent regulatory framework

by detecting existing regulation (regardless of name used)
by avoiding inconsistency re 1 plant under >1 synonyms

c) save time & effort:

where do you find ALL known synonyms?
how to maintain your taxonomy up to date?
how to compare with lists stored by other regulators?

Example 3: Regulated plant lists

In 2006, EMA supplied Kew with list of 1040 "plants under regulation" combined from lists supplied to them by FDA & Japanese Regulator

Step 1: We deduplicated the list: 1040 names => 864 unique names

Step 2: **Are these 864 names scientifically meaningful?**

Yes: perfect matches	451 [52%]
Maybe: name matched but error in author string	320 [37%]
No:	93 [11%]

Step 3: **How many are current accepted names?**

Of 451 "good" names, 209 were found in our Checklist (V2006)
of these

Accepted names	179 [85%]
Synonyms	30 [15%]

Example 3: Regulated plant lists (cont)

Step 4: **How many plants included twice?**

Japanese list had 3 plants included twice (under different names)

FDA list included 1 plant twice (::)

Potential for contradictory controls for same plant

Step 5: **How many names known for these plants?**

The list included 205 unique plants (209 – 4)

179 were referred to by their accepted name

26 were referred to by one of their synonyms

World Checklist (V.2006) recorded 2,317 synonyms for these 205 plants

i.e. you need to search Google using 2,522 names to find all published information!

4) Consequences for terminological standards

a) Plant: Name relationship many:many
b) "Author" field is critical to disambiguation:
   "*Vicia faba* L." is a different name from "*Vicia faba* Smith"
   "Plant Name Authors" are inconsistently abbreviated
c) Citation: where name used, by who & how critical:
   – particularly true for common names!!
d) Need for links to curated terminological controls:
   via #identifiers
   Existing name standards
   Author abbreviations
   Current peer-reviewed taxonomic hierarchy

Structure of Presentation

1 Introduce MPNS

2 What is a scientific plant name?

3 Issues when using scientific plant names

4 Impact & relevance

5 Global plant name resources

6 Role of MPNS and relevance to IDMP

Global Plant Name Resources

1. International Plant Name Index www.ipni.org
   - "Register" of formally published plant names: standard
   - Kew + Harvard + Australian Govt
   - complete: 1.6 million records
   - supplies, resolves identifiers (versioned)
   - incl. standard plant name author abbreviations
   - **actively curated [> 350K edits annually]**

   Limitations
   - No synonymy
   - No distribution
   - Designed for taxonomists
   - Weaker on names of varieties and infraspecific

Global Plant Name Resources

2. World Checklist of Plants www.kew.org/wcsp

- 'Gold standard' Monographs: peer reviewed accounts
- SYNONYMY for ALL species (and names) Family/Family
- Kew & network >120 taxonomic specialists/partners
- includes geography, habit
- alternative taxonomic views
- 45% families completed, peer reviewed & published online
  - Palms, Orchids, Grasses, Rubiaceae etc all available
  - Supply this data to Catalogue Life/ ThePlantList/ GBIF/ EOL etc
- **actively curated [> 120K record edits/yr]**

Limitations

- 55% families work in progress – not yet peer reviewed/published
- Website designed for botanists
- No API

Global Plant Name Resources

3. The Plant List (TPL) www.theplantlist.org

UN treaty: "Global Strategy for Conservation of Plants (GSPC)

2010 Target #1: "A working list of plants of the world"

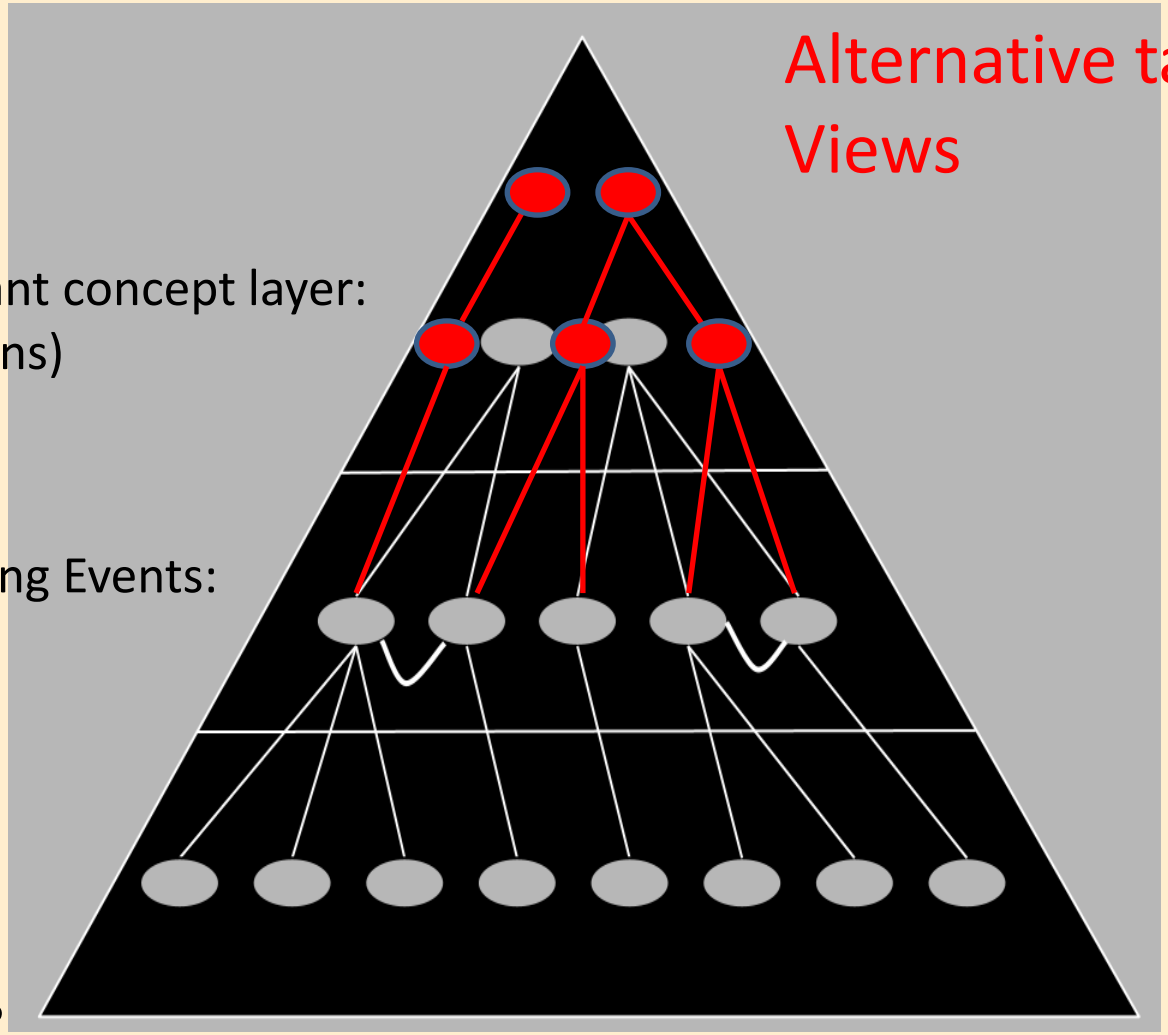
- Synthetic list: amalgamating the best of existing data sets
- ALL plant species (incl. ferns, mosses, conifers etc)
- ALL their names (1.2 million)
- Kew + Missouri Botanical Garden + partners
- Inputs from existing global family treatments AND regional Floras:
  - WCS (published /unpublished)
  - ILDIS + Rosaceae + Asteraceae + etc
  - MBG:TROPICOS: > 30 National / Regional Floras: China/Mesoamerica/NorthAmerica/ etc
- Logical rulebase: detecting / resolving conflicting opinions
- Linkages with contributors' data sets and IPNI

Limitations

- Variable quality: reflecting data provider & conflict resolution algorithms
- Not curated: Fungi & Algae yet to be added.
- Release 1 had 25% of names "unresolved" – provisional acceptance

Release 2 available March 2013: Planning a dynamic version

Kew's plant names architecture



Alternative taxonomic Views

e.g. WCS or TPL

Taxonomic / Plant concept layer:
SPECIES (opinions)

Formal Publishing Events:
NAMES (facts)

IPNI

Recorded uses of names:
OCCURRENCES
incl. misspellings, misapplied names etc

e.g. NCBI, Pharmacopoeia, legislation or journal article etc.

Structure of Presentation

1 Introduce MPNS

2 What is a scientific plant name?

3 Issues when using scientific plant names

4 Impact & relevance

5 Global plant name resources

6 Role of MPNS and relevance to IDMP

MPNS development: 2012-2015

MPNS Resources

- Plant name ontologies: IPNI, WCS, TPL
- Domain knowledge from Pharmacopoeia etc
- Links: partner resources
- Citation resource

MPNS Services

- Name Portal
- Name validation service
- Data download/curation
- API: Machine Interface
- Best Practice Guides
- Consultancies

Business Deliverables

- Project Planning
- Sustainability Planning
- Dissemination materials
- Active User Group
- Website

MPNS Stakeholders

Partners | Users | Kew/Botanists

Additional resources

Reviewed plants covered by legislation (or pharmacopoeia) and existing data resources. As a result we are:

1. Enhancing Kew's core plant name ontologies (WCS & IPNI) to achieve "Gold" standard for all medicinal plants
   - 2K plants to date

3. Adding common names, pharmacopoeia names, herbal ingredients, uses and plant parts
   - 6K plants to date

4. Linking to, & validating data resources held by partners:
   - Enrich each partner's resources
   - Facilitate maintenance of data longer term

EXAMPLE 4: "Black Cohosh": **'Cimicifugae Rhizoma'**

	British Herbal Pharmacopoeia (1996)	Chinese Pharmacopoeia (2010)	Japanese Pharmacopoeia (2006)
Common name	**Black Cohosh**	Largetrifoliolious Bugbane Rhizome	Cimicifuga Rhizome
Latin Pharmaceutical Name	**Cimicifugae Rhizoma**	**Cimicifugae Rhizoma**	**Cimicifugae Rhizoma**
Latin Scientific name(s) used in the pharmacopoeia	*Cimicifuga racemosa* (L.) Nutt.	*Cimicifuga foetida* L. or *C. heracleifolia* Kom. or *C. dahurica* (Turcz.) Maxim.	*Cimicifuga foetida* L. or *C. heracleifolia* Kom. or *C. dahurica* (Turcz.) Maxim. or *C. simplex* Wormsk.
Other name(s)	Cimicifuga (synonym)	Shengma (pin yin)	

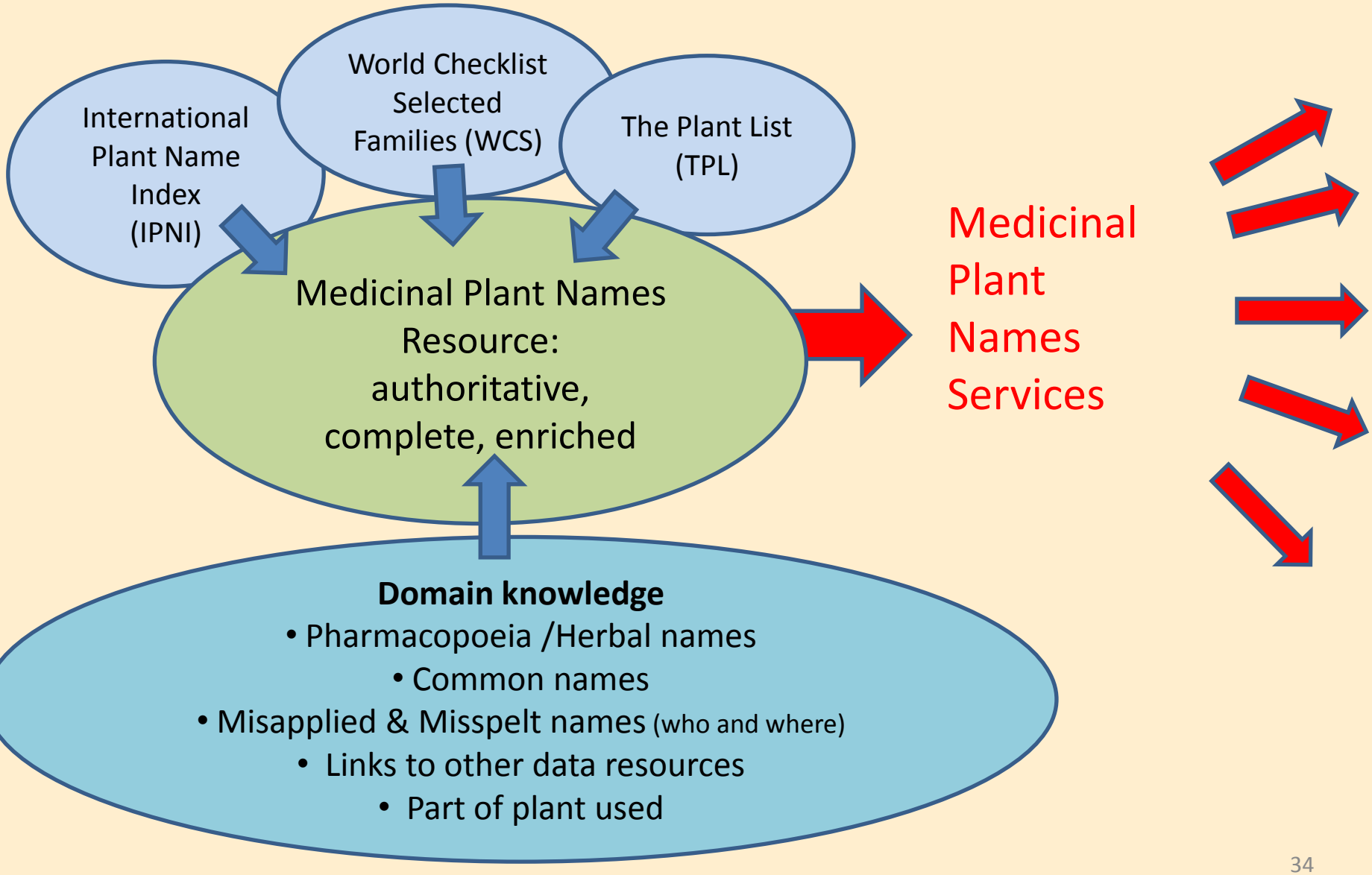
EXAMPLE 4: "Black Cohosh": **'Cimicifugae Rhizoma'**

	British Herbal Pharmacopoeia (1996)	Chinese Pharmacopoeia (2010)	Japanese Pharmacopoeia (2006)
Common name	**Black Cohosh**	Largetrifoliolious Bugbane Rhizome	Cimicifuga Rhizome
Latin Pharmaceutical Name	**Cimicifugae Rhizoma**	**Cimicifugae Rhizoma**	**Cimicifugae Rhizoma**
Latin Scientific name(s) used in the pharmacopoeia	*Cimicifuga racemosa* (L.) Nutt.	*Cimicifuga foetida* L. or *C. heracleifolia* Kom. or *C. dahurica* (Turcz.) Maxim.	*Cimicifuga foetida* L. or *C. heracleifolia* Kom. or *C. dahurica* (Turcz.) Maxim. or *C. simplex* Wormsk.
Other name(s)	Cimicifuga (synonym)	Shengma (pin yin)	
Use(s)	Menopause, rheumatism	Headache, toothache, diarrhoea, measles etc	Headache, toothache, diarrhoea, measles etc

'Cimicifugae Rhizoma' may refer to FIVE different plant species!

Actaea racemosa L.	YES		
A. cimicifuga L.		YES	YES
A. heracleifolia *(Kom.) J.Compton*		YES	YES
A. dahurica *(Turcz. ex Fisch. & C.A.Mey) Franch.*		YES	YES
A. simplex *Wormsk. Ex Fisch. & C.A.Mey*			YES

Medicinal Plant Names Services



Services

1. **Portal**
   - Pages: “Plants”, “Herbal ingredients”(drug); “Reference sources”
   - Search by : Scientific, Common or Pharmacopoeia name
   - Name resolution
     - validation (incl. fuzzy matching)
     - disambiguation / pointers to alternative uses/meanings
     - embedded links to authoritative source
   - Browse by:
     - Plant Taxonomy
     - Drug Taxonomy/Source

2. **Plant name validation service**
   - batch cleaning and enrichment of client data sets
   - sophisticated name matching algorithms (incl. fuzzy and context rules)
   - conflict resolution rules build on rules employed in “ThePlantList”
   - precision improves as more data sets are processed

Services

3. Data downloads
4. API services
   - Name validation
   - Data supply
   - Enhanced Google
   - Plant name validation e.g. for Journal editors
5. Consultancy re medicinal plant names
   - Best practice guidelines (workflows)
   - Systems design
   - Authentication / DNA/ Chemical profiling

Example partnerships/collaborations

- WHO
- MHRA
- British and Chinese Pharmacopoeia
- Medicinal and Aromatic Plants Resource db
- American Herbal Pharmacopoeia
- IUCN: medicinal plant specialist group (conservation)
- British Standards Institute
- UK National Poisons Information Service
- EMA
- UCL School of Pharmacy
- Journal of Ethnopharmacology (Elsevier)
- Pharmaceutical Press
- NAPRALERT?

Relevance of MPNS to IDPM?

A) Core botanical terminologies

1. Standard terminological control for Latin scientific plant names: **is this a name?**
   - **all** names/plants required (whether medicinal or not)
   - the standard registry of these names as they are published
2. Comprehensive, authoritative name ontologies: **which are synonyms? which belong to same genus?**
   - ALL medicinal plant names covered
   - alternative taxonomies referenced
   - global network of botanical expertise: quality
3. Actively curated

Relevance of MPNS to IDPM?

B) Improving access and relevance

4. Enrichment of core resources:
   addressing needs of health/medicinal audience
   - map common names, pharmacopoeia names, uses, plant part
   - resolve scientific names actually used – **and misused** - within the domain
   - link to relevant resources
5. Services tailored to those needs via User Group
   - validate and enrich plant names in other people's data
   - Portal: search by ANY class of name
   - API
   - advice/guidance re data structures/workflows
6. Long-term
   - RBG Kew & community commitment to long term curation core data
   - user group
   - partnerships: sharing costs/mutual benefits

Thanks for your attention

International Plant Name Index: www.ipni.org
World Checklist: www.kew.org/wcsp
The Plant List: www.theplantlist.org

… as from next month
Medicinal Plant Name Services: www.kew.org/mpns

Bob Allkin, Information Products Manager
Royal Botanic Gardens, Kew, UK
b.allkin@kew.org